

Probability Review

The toolbox beneath the course — and where each tool was used

Jinkyoo Park · KAIST

A handful of facts about *expectation*, *Bayes* and the *Gaussian*, carrying an entire course

— ORIENTATION

Orientation

Use this as a refresher before a lecture or return to it when a probability step moves too quickly. Each topic points to where it is used.

Learning route — four tools to calculate, not memorize

Bring: Algebra, sums and the notation for a conditional probability.

FIRST PASS	WHAT TO DO
Follow the idea	Expectation → conditioning → Gaussian calculations → sampling
Work without the solution	Decompose predictive variance and identify which uncertainty more measurements can reduce.
Return later	Use this chapter when a calculation in Lectures 2–12 needs a probability identity.

For the temperature thread: **predict** → **calculate** → **reveal and check** → **change one condition**. Complete the core calculation before reading the research extensions.

— EXPECTATION

Expectation and its laws

Almost every objective in this course is an expectation. Two laws about *conditioning* do most of the work.

Mean, variance, and the mean of a function

For a random variable U with density p , and any $g : \mathbb{R} \rightarrow \mathbb{R}$:

$$\mathbb{E}[U] = \int u p(u) du, \quad \mathbb{E}[g(U)] = \int g(u) p(u) du, \quad \text{var}(U) = \mathbb{E}[(U - \mathbb{E}[U])^2] = \mathbb{E}[U^2] - \mathbb{E}[U]^2$$

Linearity is the property used most often, and the one most often used without noticing:

$$\mathbb{E}[aU + bV] = a \mathbb{E}[U] + b \mathbb{E}[V] \quad \text{regardless of any dependence between } U \text{ and } V.$$

It is why a return can be split term by term, why subtracting a baseline can be separated algebraically (zero bias additionally needs the score identity), and why the covariance of AY is $A\Sigma A^\top$. Many probabilistic objectives in the course are expectations: the return $\mathbb{E}_\pi[\sum_t \gamma^t r_t]$ (Ch 7–11), expected improvement $\mathbb{E}[\max(0, f - f^{\max})]$ (Ch 4), the ELBO $\mathbb{E}_q[\log p(x | z)]$ (Ch 6), the policy objective $\mathbb{E}_\tau[R(\tau)]$ (Ch 10). **Manipulating expectations is the technical core of the course.**

The two laws that recur

LAW OF TOTAL EXPECTATION

average the conditional averages

$$\mathbb{E}[U] = \mathbb{E}_V[\mathbb{E}[U | V]]$$

LAW OF TOTAL VARIANCE

uncertainty splits in two, and the split is exact

$$\text{var}(U) = \underbrace{\mathbb{E}_V[\text{var}(U | V)]}_{\text{within — noise left inside each condition}} + \underbrace{\text{var}(\mathbb{E}[U | V])}_{\text{between — spread across conditions}}$$

Both say the same thing from two heights: conditional means average back to the original mean. Conditional variance is smaller **on average over** V , not necessarily for every observed value. Everything on the next two slides is those two sentences, cashed.

Total variance with numbers — within groups and between groups

Two equally likely groups have means **1 and 3**, and both have variance **1**.

$$\mathbb{E}[U] = 0.5(1) + 0.5(3) = 2.$$

CONTRIBUTION	CALCULATION	VALUE
within-group variance	$0.5(1) + 0.5(1)$	1
variance of group means	$0.5(1 - 2)^2 + 0.5(3 - 2)^2$	1
total variance	within + between	2

Learning the group removes the uncertainty about **which mean applies**. There is still variation inside that group. This is the same decomposition used in posterior prediction.

Temperature thread — separate temperature uncertainty from sensor noise

shared teaching example · predict before revealing the calculation

Given data D , the unknown temperature has posterior mean 21.6 and variance 0.8. A new reading is $Y = \theta + \epsilon$, with independent zero-mean sensor noise of variance 1.

Predict: Which variance belongs to the temperature, and which belongs to the next sensor reading?

Calculate and check. $\mathbb{E}[Y \mid D] = \mathbf{21.6}$. Total variance gives $\text{Var}(Y \mid D) = \mathbb{E}[\text{Var}(Y \mid \theta, D) \mid D] + \text{Var}(\mathbb{E}[Y \mid \theta, D] \mid D) = 1 + 0.8 = \mathbf{1.8}$.

The first term is sensor variability conditional on temperature; the second is uncertainty about the temperature itself.

Try it — the temperature becomes known exactly

work independently · reveal only after writing an answer

Keep the sensor noise model, but imagine that θ is now known exactly. What happens to the variance of a new sensor reading? Explain without treating the observation as noise-free.

Check your answer. Parameter uncertainty becomes zero, while the new-reading variance remains **1**. Knowing the temperature exactly does not eliminate the specified sensor noise.

Always name the random variable and the conditioning information before using an uncertainty statement.

Total expectation, read in three chapters

CH 2 – THE PRIOR MEAN

$$\mathbb{E}[\theta] = \mathbb{E}[\mathbb{E}[\theta \mid y]]$$

Before seeing data, your belief is already the average of every belief you might end up with. A posterior that moved in one direction *for every possible dataset* would be an incoherent prior.

CH 7 – THE BELLMAN EQUATION

$$V^\pi(s) = \mathbb{E}[r_{t+1} + \gamma V^\pi(s_{t+1}) \mid s]$$

Lecture 7 peeled one step off the return and let the Markov property close the loop. **That peel is the tower rule** — condition on the next state, average the conditional averages.

CH 8 – THE SAMPLED BACKUP

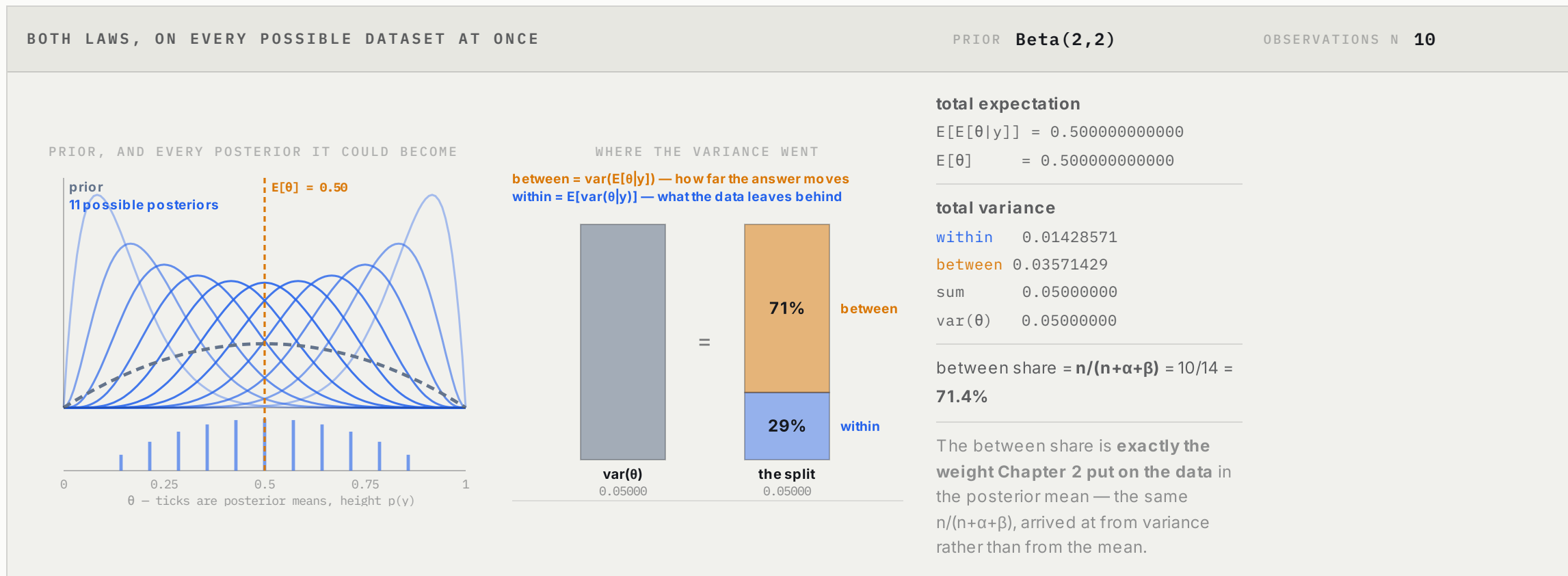
$$\mathbb{E}_{s'}[\cdot] \longrightarrow \text{one transition } (s, a, r, s')$$

Lecture 8's translation table swaps the *outer* expectation for a single draw and calls the law of large numbers to justify it. The equation is unchanged; only who evaluates it.

↔ used in Ch 2 (the prior–posterior identity), Ch 7 (every Bellman equation), Ch 8 and 10 (every sampled backup).

Total variance, on every dataset at once

the identity computed rather than asserted — every dataset enumerated, nothing sampled



A $\text{Beta}(\alpha, \beta)$ prior and n coin tosses, with *every* dataset enumerated rather than sampled. The posterior means scatter around the prior mean and average back onto it — total expectation. The variance bar splits exactly into a within part and a between part — total variance. Raise n and watch the between share grow: it is $n/(n + \alpha + \beta)$, **the very weight Chapter 2 put on the data.**

The three places that split mattered

CH 2 – THE POSTERIOR TIGHTENS

$\mathbb{E}_y[\text{var}(\theta \mid y)]$ equals prior variance minus the variance of posterior means. A particular dataset can still increase posterior variance. Chapter 2 stated it as **data cannot, on average, make you less certain** — which is this identity read as an inequality.

CH 2 AND 4 – THE PREDICTIVE SPLIT

The predictive variance is $\sigma_Y^2 + \tau_1^2$: measurement noise you can never remove, plus parameter uncertainty you can. Lecture 0's aleatoric/epistemic pair, arriving as the two terms of one identity.

CH 10 – THE BASELINE

The residual $U - \mathbb{E}[U \mid V]$ has mean **zero**, not the original mean. Policy-gradient baselines preserve the mean by a different fact: the action score has expectation zero. A well-chosen baseline can reduce variance; see Lecture 10.

Total variance explains the first two readings: *how much did I learn* and *what uncertainty remains in an observation*. For a policy-gradient baseline, also use the score identity and a variance calculation; total variance alone does not prove unbiasedness.

Check — the tower property

The law of total expectation says $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]]$. Where does this course lean on it hardest?

- ☒ A In every Bellman equation — conditioning on the next state and averaging over it is exactly this identity
- ☐ B In proving that the sample mean is unbiased
- ☐ C In deriving the Gaussian density
- ☐ D In the definition of conditional independence

Writing $V(s) = \mathbb{E}[r + \gamma V(s')]$ is the tower property with the conditioning done on the next state. Every backup in Lectures 7 to 12 is one application of it, and the reason a one-step update can stand in for an infinite sum of future rewards is that this identity says it may.

— BAYES

Bayes and the two views

One line of algebra that turns a belief and a measurement into a new belief — and two interpretations of probability and uncertainty.

Bayes' rule — the whole of inference on one line

$$p(\theta \mid \text{data}) = \frac{p(\text{data} \mid \theta) p(\theta)}{p(\text{data})} \propto \underbrace{p(\text{data} \mid \theta)}_{\text{likelihood}} \underbrace{p(\theta)}_{\text{prior}}$$

WHERE IT WAS THE WHOLE METHOD

- **Ch 2** — prior to posterior, conjugacy, and MAP as MLE plus a prior term. Regularisation turned out to *be* a prior.
- **Ch 3** — the same rule with many unknowns at once: marginalise the hidden variables, then normalise. The graph writes the joint as a product; *summing it back down* is where the cost lives.
- **Ch 4** — a prior over *functions*; the GP posterior is Bayes' rule with θ replaced by f .

WHERE IT WAS HIDING

- **Ch 6** — CbAS writes its target as $p(x \mid S) \propto P(S \mid x) p(x \mid \theta^{(0)})$: **the generative model is the prior and the oracle is the likelihood.**
- **Ch 5** — an ensemble's disagreement is a crude posterior width, used as an out-of-distribution alarm.

Belief in = prior; belief out = posterior; **the data does the turning.**

Frequentist and Bayesian — two readings of the same data

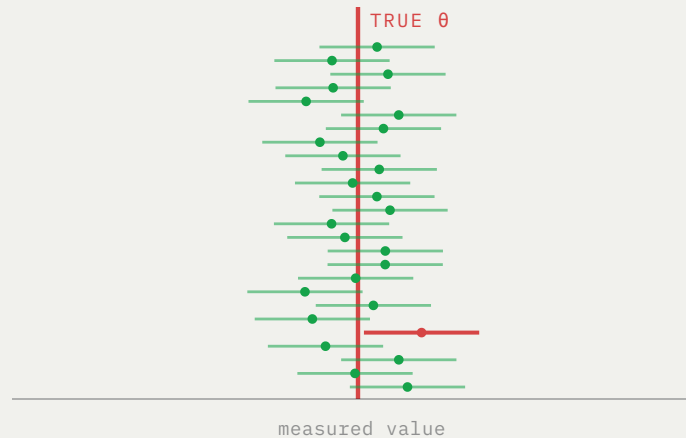
	FREQUENTIST	BAYESIAN
what is random	θ fixed, the data random	the data fixed, θ random
what varies	variability <i>in the data</i> given θ	variability <i>in θ</i> given the data
the interval	confidence: $\theta \in \text{CI}$, or not	credible: $\Pr(\theta \in \text{CR}) = 0.95$
the sentence	95% of repeated intervals cover θ	95% chance θ lies in <i>this</i> region

The distinction is not pedantry. It is the MLE-versus-MAP, point-versus-distribution choice that Chapter 2 was built on, and it is why a Gaussian process reports a credible **band** rather than a confidence *statement*.

Both say 95%, and they are different claims

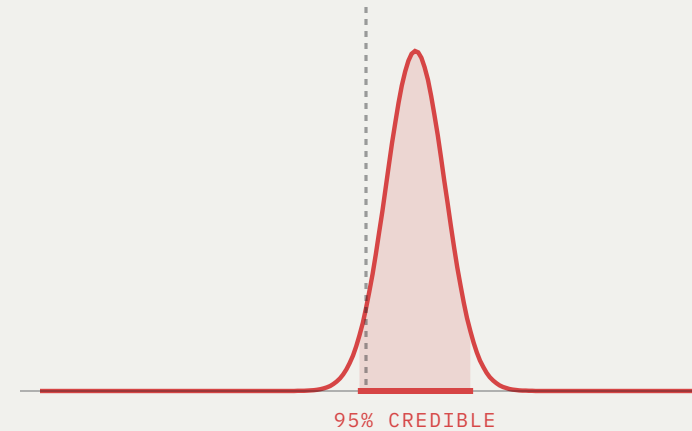
BOTH SAY 95% — AND THEY ARE DIFFERENT CLAIMS

FREQUENTIST — VARIABILITY IN THE DATA



25 of 26 intervals cover the true θ .
The 95% is a property of **the procedure**, not of any one interval.

BAYESIAN — VARIABILITY IN θ



From **this dataset**: $\Pr(\theta \in \text{CR} \mid \text{data}) \approx 0.95$.
Known Gaussian noise; flat prior on the mean.

Left, the frequentist picture: θ is a fixed vertical line and the *intervals* are what move — run the experiment 26 times and about 95% of them cover it. Right, the Bayesian picture: one dataset, and the *parameter* is what is spread. Chapter 2 ran this widget for the same reason.

The other two views — aleatoric and epistemic

ALEATORIC — IRREDUCIBLE

The world is genuinely stochastic. Repeating the experiment gives a different answer, and no amount of data removes it. All you can do is *characterise* it, as a distribution.

Measurement noise σ_Y^2 (Ch 2), observation noise σ_ϵ^2 on the GP diagonal (Ch 4), the transition kernel's own randomness (Ch 7).

EPISTEMIC — REDUCIBLE

We simply do not know the model: its parameters, its state, its dynamics. **Data shrinks it, and the whole course is a campaign against it.**

Parameter uncertainty τ_1^2 (Ch 2), the GP band away from data (Ch 4), ensemble disagreement (Ch 5), and PETS' ensemble spread far from data, whose *per-member* variance is the aleatoric half (Ch 11).

Lecture 0 named the pair on the slide *Two kinds of uncertainty — name them once*; Chapter 2's Act 4 ended by writing them as the two terms of $\sigma_Y^2 + \tau_1^2$. That is the law of total variance again — the *within* part is aleatoric, the *between* part epistemic — which is why this appendix keeps circling the same identity.

Check — the estimator's wall

A Monte Carlo estimate from n samples has standard error $\sigma / \sqrt{n}$. To halve the error, you must:

- ☐ A Double the number of samples
- ☒ B Quadruple the number of samples
- ☐ C Halve σ ; n has no effect on the error
- ☐ D Increase n by a factor of $\sqrt{2}$

The $\sqrt{n}$ is a wall, not a slope: each additional digit of accuracy costs a hundredfold more samples. This single fact is why variance reduction is worth so much effort — baselines in Lecture 10, control variates, doubly-robust estimators in Lecture 12. Shrinking σ buys what growing n can only buy at enormous cost.

— GAUSSIAN

The multivariate Gaussian — the workhorse

The single most-used distribution in the course. Four properties, and each one buys a chapter.

The multivariate normal — definition

A nonsingular jointly Gaussian vector $Y = (Y_1, \dots, Y_k)$ has density

$$p(y) = \mathcal{N}(y \mid \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \exp\left(-\frac{1}{2}(y - \mu)^\top \Sigma^{-1}(y - \mu)\right)$$

For this density require $\Sigma \succ 0$, with $\mathbb{E}[Y] = \mu$ and $\text{var}(Y) = \Sigma$. A singular Gaussian still exists but has no full-dimensional density of this form; $\Sigma_{ij} = \text{cov}(Y_i, Y_j)$.

Four properties make it the backbone of [Gaussian-process regression](#) — and of half the closed forms in this course.

Each is a one-line fact. Together they are why so much of the course can carry uncertainty *analytically* rather than by sampling: a Gaussian goes in, a Gaussian comes out, and the parameters update by matrix algebra rather than by an integral you cannot do.

Property 1— for Gaussians, uncorrelated means independent

Setting $\Sigma_{ij} = 0$ for $i \neq j$ makes Σ and Σ^{-1} diagonal, so the joint density factors:

$$\Sigma \text{ diagonal} \implies p(y) = \prod_i \mathcal{N}(y_i \mid \mu_i, \sigma_{ii})$$

For general random variables this is false. Take $X \sim \mathcal{N}(0, 1)$ and $Y = X^2$: then $\text{cov}(X, Y) = \mathbb{E}[X^3] = 0$, and yet Y is a *function* of X — as dependent as two variables can be.

Zero correlation is a statement about **second moments only**. **The Gaussian is special** because it has nothing else.

↪ every i.i.d. noise model in the course leans on this — it is what lets a diagonal Σ stand in for genuine independence.

Property 2 — a linear map of a Gaussian is Gaussian

$$Z = AY \sim \mathcal{N}(A\mu, A\Sigma A^\top)$$

Both moments are just linearity: $\mathbb{E}[AY] = A\mu$ and $\text{cov}(AY) = A\Sigma A^\top$. No integral is needed, and **the family is closed under the operation** — which is the property that makes Gaussians tractable everywhere they appear.

↪ the reparameterisation trick $z = \mu_\phi + \epsilon\sigma_\phi$ (Ch 6), and linear-Gaussian dynamics $x_{t+1} = Ax_t + Bu_t + w_t$ (Ch 9) — in both, a Gaussian is pushed through a linear map and must come out Gaussian for the method to close.

Properties 3 and 4 — marginals and conditionals stay Gaussian

Property 3 — marginals are Gaussian. Any sub-vector of a Gaussian is Gaussian; read off the matching blocks of μ and Σ and ignore the rest. (It is Property 2 with a selector matrix.)

Property 4 — conditionals are Gaussian. The one that makes GP regression work. For

$$Z = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right) \implies Y_2 \mid Y_1 = y \sim \mathcal{N}\left(\underbrace{\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(y - \mu_1)}_{\text{posterior mean}}, \underbrace{\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}}_{\text{posterior covariance}}\right)$$

Read the two braces. The mean moves by **how far the observation fell from its own mean, geared by the correlation**; with fixed covariance parameters, the covariance drops by a term independent of the observed value — which is why a GP's error bars can be planned before any data arrives, and why Chapter 4 could choose where to sample next.

Gaussian conditioning — read one observation across a correlation

Let Y_1, Y_2 have means 0, variances 1, and correlation **0.8**. Observe $Y_1 = 1$.

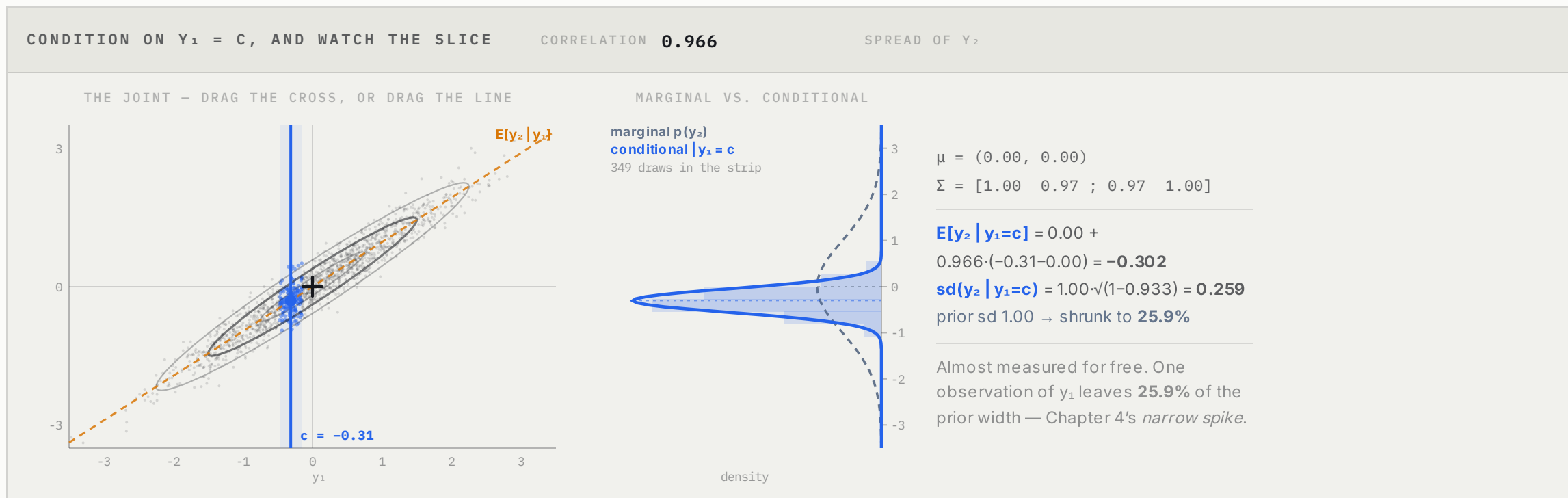
$$\mathbb{E}[Y_2 \mid Y_1 = 1] = 0 + 0.8(1) = 0.8,$$

$$\text{Var}(Y_2 \mid Y_1 = 1) = 1 - 0.8^2 = 0.36.$$

So $Y_2 \mid Y_1 = 1 \sim \mathcal{N}(0.8, 0.36)$, with standard deviation **0.6**.

Correlation transmits information: the mean shifts toward the observation, and uncertainty shrinks. If the correlation were zero, neither would change. If the observation were -1 , the mean would be -0.8 but the variance would still be 0.36.

Conditioning, in your hands



Drag the mean, reshape Σ with the correlation and scale controls, then drag the conditioning line $y_1 = c$. The blue slice is the exact conditional: its mean slides along the regression line $\mu_2 + \rho \frac{\sigma_2}{\sigma_1}(c - \mu_1)$ and its width shrinks to $\sigma_2 \sqrt{1 - \rho^2}$ — **a factor that depends on the correlation and on nothing else**. The two presets are Chapter 4's own numbers.

Where Property 4 did the work

CH 4 — THE GP POSTERIOR

Take $Y_1 = \mathbf{y}_{1:n}$ and $Y_2 = f(x)$, so $\Sigma_{11} = \mathbf{K} + \sigma_\epsilon^2 \mathbf{I}$, $\Sigma_{21} = \mathbf{k}^\top$, $\Sigma_{22} = k(x, x)$:

$$\mu(x \mid \mathcal{D}) = \mathbf{k}^\top \Sigma_{11}^{-1} \mathbf{y}.$$

A GP posterior is Gaussian conditioning, and nothing else.

CH 2 — THE WEIGHT POSTERIOR

With $y = \mathbf{X}w + \epsilon$ and both w and ϵ Gaussian, w and y are *jointly* Gaussian — so $p(w \mid \mathbf{X}, y)$ is Property 4 again.

It lands on $\mathcal{N}(\mu_w, \Sigma_w)$ with μ_w the ridge solution, $\lambda = \sigma^2 / \alpha^2$.

CH 9 — LQR WITH NOISE

For finite-horizon, fully observed LQ control with additive zero-mean noise independent of state and action, the value is quadratic **plus a constant**. Put noise back into $x_{t+1} = Ax_t + Bu_t + w_t$ and

$$x \sim \mathcal{N}(\mu, \Sigma),$$

$$\mathbb{E}[x^\top Px] = \mu^\top P\mu + \text{tr}(P\Sigma).$$

The noise adds a *constant*. The gain K is unchanged — certainty equivalence, in one line.

Chapter 4 quantified this at two correlations from its own kernel. At $\text{corr}(f_1, f_2) = 0.966$ one observation shrinks the neighbour's standard deviation to $\sqrt{1 - 0.966^2} = 0.259$ of the prior — *a narrow spike*. Four points apart, at $\text{corr} = 0.573$, it shrinks only to 0.820 — *barely narrower than the prior*, a value you still have to buy. Both are $\sqrt{1 - \rho^2}$ and nothing more.

Check — the property everything rests on

Which property of the multivariate Gaussian is used most often in this course?

- ☐ A That it maximises entropy for a given mean and covariance
- ☐ B That the sum of independent Gaussians is Gaussian
- ☒ C That conditioning on part of the vector leaves a Gaussian, with a closed form for its mean and covariance
- ☐ D That it is the limit distribution of the central limit theorem

Gaussian conditioning **is** the Gaussian process posterior, and the GP posterior is Lecture 4's entire engine and Lecture 5's surrogate. Marginals stay Gaussian, conditionals stay Gaussian, and linear maps of Gaussians stay Gaussian — which together are why so much of this course has closed-form answers instead of sampling loops.

— SAMPLING

Sampling and divergences

What to do when the integral has no closed form — and how to say that one distribution is far from another.

Monte Carlo — when the integral is intractable

When an expectation has no closed form, **estimate it by sampling**:

$$\mathbb{E}_{x \sim p}[g(x)] \approx \hat{\mu}_N = \frac{1}{N} \sum_{i=1}^N g(x_i), \quad x_i \sim p$$

For i.i.d. samples and an integrable g , the average is unbiased. If variance is finite, its exact standard error is $\sigma_g/\sqrt{N}$; the central limit theorem supplies an asymptotic Gaussian approximation. Correlated samples require a different standard-error calculation.

This is the quiet foundation of *model-free* reinforcement learning. When the expectation over unknown dynamics cannot be computed, average over sampled trajectories instead: Monte-Carlo returns and TD targets (Ch 8), REINFORCE and advantage estimates (Ch 10), and the ELBO's reconstruction term, which the reparameterisation trick turns into a one-sample average (Ch 6). ↪ *Ch 8 and 10 — the entire data-driven half of the course.*

For an i.i.d. average the rate $1/\sqrt{N}$ is fixed by the central limit theorem. **Variance-reduction methods can lower the constant, subject to cost and assumptions.**

Importance sampling — sample here, answer about there

$$\mathbb{E}_{x \sim p}[g(x)] = \mathbb{E}_{x \sim q} \left[\underbrace{\frac{p(x)}{q(x)}}_{\text{weight } w(x)} g(x) \right] \quad \text{for any } q \text{ with } q > 0 \text{ wherever } pg \neq 0$$

WHERE IT WORKED — CH 6

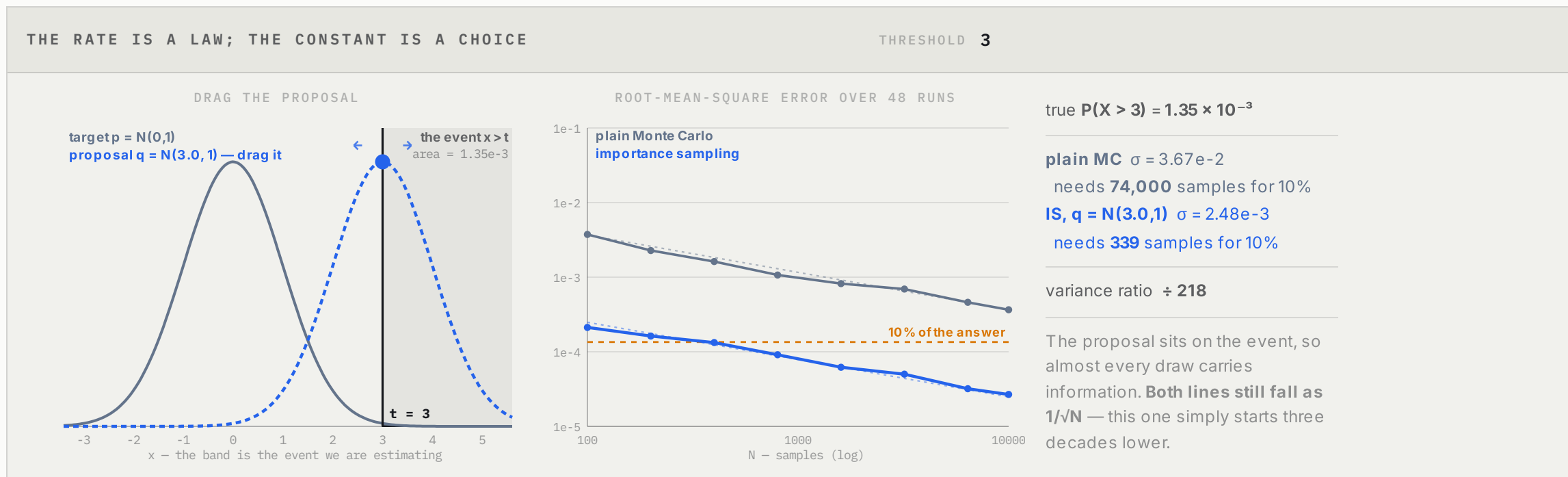
CbAS cannot draw from $p(x \mid S)$ directly, so it draws from the previous iterate and reweights. The ladder exists precisely to keep q close to the target, **which helps control weight variability but does not guarantee bounded weights or finite variance.**

WHERE IT IS THE WHOLE PROBLEM — CH 12

Off-policy evaluation scores a policy you may not deploy, by reweighting a dataset another policy collected. The weight is a *product over the horizon*, so its variance grows with T — which is why the chapter reaches for doubly-robust estimators.

Unbiased for *any* legal q . The catch is entirely in the variance: $\text{var}_q(wg)$ can be far smaller than $\text{var}_p(g)$, far larger, or infinite — and unbiasedness will not warn you which.

The rate is fixed; the proposal sets the constant



Estimating $\Pr(X > 3) = 1.35 \times 10^{-3}$ for $X \sim \mathcal{N}(0,1)$. Both clouds of points fall as $1/\sqrt{N}$ and sit on their own exact $\sigma/\sqrt{N}$ line: the rate is not negotiable. What moves is the constant — a proposal aimed at the event needs **339 samples for 10% accuracy, plain sampling needs 74,000, and one aimed the wrong way needs 3.4 million** and returns a confident zero until it does not.

KL divergence and entropy — comparing distributions

$$\text{KL}(p \parallel q) = \mathbb{E}_{x \sim p} \left[\log \frac{p(x)}{q(x)} \right] \geq 0, \quad = 0 \iff p = q$$

Not symmetric, and not a metric — but the natural *information cost* of using q where p is true. For Gaussians it is closed form, which is why it can sit inside a loss: $\text{KL}(\mathcal{N}(0, 1) \parallel \mathcal{N}(0, 4)) = 0.318$ while the reverse is 0.807.

- **Ch 6, the VAE.** The ELBO's KL term pulls the encoder onto the prior, which is what makes the latent space **samplable**: without it q_ϕ is free to encode anywhere and new draws decode to nothing.
- **Ch 6, generative training.** Minimising KL to the data distribution *is* maximum likelihood; minimising JS instead is the GAN. Same shape, different divergence, different failure mode.
- **Ch 10, trust regions.** TRPO maximises an importance-weighted advantage subject to $\mathbb{E}[\text{KL}(\pi_{\text{old}} \parallel \pi_\theta)] \leq \delta$ — the largest step that keeps the new policy *close in behaviour*, not close in parameters. PPO discourages some large changes with clipping; it imposes no hard KL bound.

Entropy $H(p) = -\mathbb{E}_p[\log p(x)]$ measures spread rather than distance. Maximising it keeps a policy exploratory — the seed of maximum-entropy RL — and it is the term that turns a hard argmax into a soft, probabilistic best response.

Check — which divergence, and why it is asymmetric

The KL divergence $\text{KL}(q||p)$ is not symmetric. In variational inference, minimising it over q produces what behaviour?

- ☐ A q spreads out to cover every mode of p
- ☐ B The asymmetry has no practical consequence for the fitted q
- ☐ C q matches the mean and variance of p exactly
- ☒ D q is penalised heavily for putting mass where p has none, so it tends to fit a single mode and under-cover

The integrand carries $q \log(q/p)$: wherever q is large and p is near zero the penalty explodes, but where p is large and q near zero it costs almost nothing. So this direction is **mode-seeking** — it would rather explain part of the distribution well than all of it badly. This tendency depends on a restricted approximating family; a flexible family can match the whole posterior. It is not a general explanation of blurry VAE reconstructions, which also depend on the decoder likelihood and capacity.

— CLOSING

Closing

Everything above, on one page.

The toolbox map — one glance

PROBABILITY TOOL	LECTURES THAT DEPEND ON IT
expectation, linearity	all of Part IV — returns and values
law of total expectation	the Bellman equation (Ch 7–11)
law of total variance	the predictive split (Ch 2, 4); baselines and advantage (Ch 10)
Bayes' rule	Ch 2, 3, 4; CbAS's target (Ch 6)
Gaussian, Property 4	Ch 4 — the GP posterior <i>is</i> this formula
Gaussian, linear map	Ch 6 (reparameterisation), Ch 9 (LQR noise)
Monte Carlo estimation	Ch 8, 10 — model-free RL
importance sampling	Ch 6 (CbAS), Ch 12 (off-policy evaluation)
KL, entropy	Ch 6 (VAE), Ch 10 (TRPO)

Two objects, and the rest is bookkeeping

The whole course manipulates an **expectation** — what we believe will happen on average — and a **Gaussian** — one especially tractable uncertainty model.

Everything else is how to **update** them (Bayes, conjugacy, conditioning), how to **estimate** them when the integral will not close (Monte Carlo, importance sampling), or how to **bound the gap** between two of them (KL, trust regions).

Keep this appendix beside the main lectures. Whenever a derivation moves quickly through a probability step — a swapped expectation, a conditional that stays Gaussian, a sample average standing in for an integral — the justification is one of the nine rows above, and the pointer beside it says where to look.

An expectation, and a Gaussian.

Update them, estimate them, or bound the gap between them — that is the course.

— APPENDIX

Examples, readings and derivations

Reference material for a second pass. The main teaching route ends before this section.

Reading guide — read the tool where the course first needs it

one main idea to explain, one comparison, one application

ROLE	READ OR REVISIT	QUESTION TO ANSWER
Core	Murphy, <i>Probabilistic Machine Learning: An Introduction</i> (2022): probability and Gaussian sections	Which random variable and conditioning information appear in each expectation?
Compare	Sutton & Barto, <i>Reinforcement Learning: An Introduction</i> (2018): returns and sampled estimates	What changes when an expectation is approximated from trajectories?
Apply	Coin prediction, GP conditioning and policy-gradient baselines	Which identity justifies the calculation, and what assumptions does it need?

This is a reference toolbox, not a survey of new papers. Work the small examples and return to the relevant lecture; no separate research-paper assignment is required.